**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2017/M0608**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of March 14, 2017**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2017\_03\_14 15\_07.rtf**

**Robert Reyling (to Everyone)**: 1:15 PM: Do we need a 'Mission statement' for this Working Group?

**Wo Chang (to Everyone)**: 1:16 PM: Thanks Robert for your comment, yes, we do have charter/scope statement for NBD-PWG as well as for each subgroup.

**Tim Zimmerlin (to Everyone)**: 1:18 PM: Frank, we have addressed use of the NBD-PWB standards. Use cases. Existing stacks like OpenStack, AWS, Azure, GCE. Applications like Hadoop and Spark and Google TensorFlow.

**Tim Zimmerlin (to Everyone)**: 1:19 PM: ...controlled vocabularies for HW and SW and distributed algorithms.

**Tim Zimmerlin (to Everyone)**: 1:20 PM: ...environmental constraints and interfaces including clouds, SQL, newSQL, local direct attached storage, GPUs, FPGAs, ASICs.

**Frank Farance (to Everyone)**: 1:21 PM: Tim, yes I agree. But use-cases are inputs into the process to inform us. So how would with the docs AFFECT us. Let's say I use Hadoop/Spark, and I read the NIST docs (or even the ISO docs) ... now what? How do these docs have an effect upon my systems?

**Tim Zimmerlin (to Everyone)**: 1:22 PM: ...laws and policies including PII, logging & auditing, national borders, notification of spills, data retention periods.

**Tim Zimmerlin (to Everyone)**: 1:23 PM: Franks, "standards are great...everyone has their favorites."

**Tim Zimmerlin (to Everyone)**: 1:23 PM: ...global marketplaces will decide in time...we will be leaders.

**Tim Zimmerlin (to Everyone)**: 1:26 PM: ...FedRAMP, AWS, Azure.

**Frank Farance (to Everyone)**: 1:31 PM: Frank Farance, +1 917 751 2900, frank@farance.com

**Frank Farance (to Everyone)**: 1:33 PM: Tim: even if there are mulitple standards, ours (NIST BDWG) should have some understanding of how this applies to (affects) existing systems, services, organizations, and data.

**Tim Zimmerlin (to Everyone)**: 1:34 PM: Frank, the NBD-PWG is NOT a normative standards body. We are analysts. I point you to IEEE as a normative standards body NIST is feeding into.

**Sumit Shah (to Everyone)**: 1:35 PM: @Arnab @Wo I would appreciate if you can share this document.

**Arnab Roy (to Everyone)**: 1:38 PM: @Sumit: sure. I believe it's already online or will be soon

**Tim Zimmerlin (to Everyone)**: 1:38 PM: Frank, the National Academies is NOT a normative standards body either. They are very active in data science.

**Sumit Shah (to Everyone)**: 1:38 PM: @Arnab - Thanks. I will look for it.

**Frank Farance (to Everyone)**: 1:39 PM: Tim: Actually, NIST is an Accredited Standards Development Organization (SDO). But even it it weren't, the NIST multi-part series is, in fact, a normative document.

**Tim Zimmerlin (to Everyone)**: 1:40 PM: Frank, the NBD-PWG is NOT NIST or even authoritative.

**Frank Farance (to Everyone)**: 1:40 PM: Just like internal software development company documents are, in fact, normative documents.

**Tim Zimmerlin (to Everyone)**: 1:41 PM: ...NIST is a Federal leader in Cloud Computing, Cyber Security, IoT, CPS.

**Frank Farance (to Everyone)**: 1:42 PM: That doesn't rebut my points that (1) NIST is an accredited SDO, and (2) NIST produces normative documents.

**Tim Zimmerlin (to Everyone)**: 1:42 PM: ...we are not NIST...

**Tim Zimmerlin (to Everyone)**: 1:43 PM: ...check the NIST Web portal for a raft of NIST standards relevant to big data.

**Frank Farance (to Everyone)**: 1:43 PM: We are participants in a NIST working group.

**Tim Zimmerlin (to Everyone)**: 1:46 PM: ...please read the caveats in the front material of the 9 volumes.

**Frank Farance (to Everyone)**: 1:51 PM: Tim, yes, I read the caveats. A normative document (or not) is determined by its content, not by its author, nor by its caveats. Take a look at NIST SP 1500-6, which is the first edition of our Reference Architecture. Clearly it is a normative document.

**Frank Farance (to Everyone)**: 1:53 PM: Here is the first sentence of Clause 5, of the NBDRA: "The Big Data characteristics of volume, velocity, variety, and variability demand a versatile management platform for storing, processing, and managing complex data. Management of Big Data systems should handle both system- and data-related aspects of the Big Data environment". Clearly, that sentence is a provision (of variety "recommendation"), and there are many many provisions within the document series. These are normative documents.

**Arnab Roy (to Everyone)**: 1:54 PM: Thanks everyone for the great discussions. I will have to leave now.

**Frank Farance (to Everyone)**: 1:54 PM: Arnab, send me your contact info.

**Arnab Roy (to Everyone)**: 1:55 PM: @Frank: I sent you an email just now

**Arnab Roy (to Everyone)**: 1:55 PM: In case you don't get it: aroy@us.fujitsu.com

**Tim Zimmerlin (to Everyone)**: 1:58 PM: Gregor, I recommend you follow this link: http://www.rle.mit.edu/ncrc/ Professor Muriel Medard has very insightful talks of networked, computable encodings. There is an algebra of encodings that simultaneously optimize redundancy / resilience in noisy channels, compression, and native computability across restricted sets on functions.

**Tim Zimmerlin (to Everyone)**: 1:59 PM: ...Lambda computing (aka microservices) is optimized in this manner.

**Tim Zimmerlin (to Everyone)**: 2:01 PM: ...your array of virtualized "X" (e.g., VMs, Virtual File Systems, Virtual SQL) need NOT be installed in the traditional sense of VMware hypervisors & services.

**Mark Underwood (to Everyone)**: 2:13 PM: This piece by Adrian Cockroft reminds us that Netflix did not go down, and is resilience-ready https://read.acloud.guru/evolution-of-business-logic-from-monoliths-through-microservices-to-functions-ff464b95a44d#.y041s3269

**Mark Underwood (to Everyone)**: 2:14 PM: He writes "In other words, it’s possible to send 100 to 1000 messages between services in the same amount of time as communicating and processing one message would take a decade ago. This is a key enabler for the move away from monolithic applications."

**Tim Zimmerlin (to Everyone)**: 2:16 PM: Frank, whose ALU? What does an ALU accomplish?

**Frank Farance (to Everyone)**: 2:21 PM: See <https://en.wikipedia.org/wiki/Arithmetic\_logic\_unit>

**Frank Farance (to Everyone)**: 2:22 PM: I didn't find any references to a new understanding of ALU, as you describe.

**Tim Zimmerlin (to Everyone)**: 2:23 PM: Frank, it is in the common literature. I have referenced Lamdba, microservices, 2 phase Lambda processing, eventually consistent distributed algorithms.

**Tim Zimmerlin (to Everyone)**: 2:24 PM: ...just re read my emails from 2013 on.

**Tim Zimmerlin (to Everyone)**: 2:24 PM: ...a good starting point is the above MIT link.

**Tim Zimmerlin (to Everyone)**: 2:25 PM: Frank, your ALU Wiki reference is OBE...GPUs, OpenCL, MPI, FPGAs, ASICs.

**Frank Farance (to Everyone)**: 2:33 PM: Tim, it is YOU who spoke about ALUs, and I was following up. I doin't see any rerefences on this. The notion of an "ALU" in the context of lambda processing, is a tiny granual component. See the Wikipedia article on lambda architecture. As an analogy, it's like speaking about Mass Transportation and a wheel's hub caps in the same sentence, two very different levels of granularity

**Tim Zimmerlin (to Everyone)**: 2:36 PM: Frank, I was giving you credit...when I used ALU now. The future of exascale pivots on future ALUs (not my term but yours).

**Frank Farance (to Everyone)**: 2:37 PM: Did you just use ALU about 60 seconds ago in the conversation with Gregor?

**Tim Zimmerlin (to Everyone)**: 2:37 PM: ...I am working on both "hard" and "soft" cores including ARM Mx, RISC V, OpenPower, x86 (both 32 bit & 64 bit).

**Frank Farance (to Everyone)**: 2:41 PM: And the notion of ALU is the same in those cores as it is in the wikipedia article on ALUs.

**Tim Zimmerlin (to Everyone)**: 2:43 PM: ...ALU has been changing since the 1950s. Cray at CDC had one CP plus 4 or more PPs. Vector machines. BCD, 1s complement, 2s complement, IEEE numbers, DEC CAR/CDR. etc.

**Tim Zimmerlin (to Everyone)**: 2:45 PM: ...Lisp was created based on the DEC ALU.

**Tim Zimmerlin (to Everyone)**: 2:47 PM: ...Lisp is the "original" Lambda environments ala Church.

**Frank Farance (to Everyone)**: 2:47 PM: Whil;e implemenmtations might change, they still correspond to the same "notion". Here is your Power8 architecture, and they refer to an ALU "an arithmetic and logic unit (ALU) to execute add, subtract, compares and

trap instructions", see page 94 of https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjpj5qu09bSAhWh24MKHYCuB0oQFggcMAA&url=https%3A%2F%2Fwww.ibm.com%2Fdeveloperworks%2Fcommunity%2Fwikis%2Fform%2Fanonymous%2Fapi%2Fwiki%2F61ad9cf2-c6a3-4d2c-b779-61ff0266d32a%2Fpage%2F1cb956e8-4160-4bea-a956-e51490c2b920%2Fattachment%2F5d3361eb-3008-4347-bf2f-6bf52e13f060%2Fmedia%2FThe%2520Power8%2520Core%2520MicroArchitecture%2520earlj%2520V5.0%2520Feb18-2016VUG2.pdf&usg=AFQjCNElfE6jKEtw0gTDyZ4jKK79XH3\_rw&sig2=\_aAinpfbaZjODu2tY59gJQ

**Frank Farance (to Everyone)**: 2:48 PM: Tim, I'm just saying that what you're saying with repsect to ALUs is not consistent with industry literature and practice.

**Tim Zimmerlin (to Everyone)**: 2:49 PM: Frank, ok. But you ignore GPUs, ASICs, larger scale buses and networks. A foreground vs background de emphasis.

**Tim Zimmerlin (to Everyone)**: 2:50 PM: Frank, research Microsofts's Remote DMA FPGAs, as one example, a starting point.

**Frank Farance (to Everyone)**: 2:51 PM: I didn't ignore them, they are different notions (concepts) of which some of them borrow features (or not) from ALUs. GPUs certainly take features of ALUs. ASICs \*can\* take features of ALUs, assuming you're dropping an ALU into an ASIC (or FPGA).

**Tim Zimmerlin (to Everyone)**: 2:51 PM: ...routers & switches have special purpose ALUs, CAM, and HW accelerators.

**Tim Zimmerlin (to Everyone)**: 2:51 PM: ...shift registers, frame buffers, error detect & correction functions.

**Frank Farance (to Everyone)**: 2:52 PM: Right, but both data/management planes have ALUs (possibly optimized differently). And shift registers, etc., are all cases of using, limiting, or extending ALUs.

**Tim Zimmerlin (to Everyone)**: 2:53 PM: ...you win...I give up!

**Cavan (Private)**: 2:57 PM: when I put togeter an updated bio, I decided to update (I never had) my Linked in with some of my publications..... So it should I should get it to you today. (Doing some adminis-trivia right now.. due)

**Ann Racuya-Robbins (to Everyone)**: 3:07 PM: Thank you all.

**Sumit Shah (to Everyone)**: 3:07 PM: thank you